

日本語対話処理のための ユーザーモデル獲得・同定手法の検討

鷲崎 誠司 西山 敏雄

NTT情報通信処理研究所

1. まえがき

自然語入力による対話処理システムにおいては、ユーザーモデルの利用により利用者対応の適切な対話を実現することが重要な要素となる。

本稿では、自然語入力文中に存在する様々な意味属性を含む形態素により入力文が表現するユーザーモデルを定義し、この定義に基づき形態素間の意味的関係の自動獲得を行うユーザーモデルの自動獲得手法、及び獲得・蓄積されているユーザーモデルの集合から新規入力文が表現する未知のユーザーモデルを、蓄積された形態素間の意味的関係の類似照合により同定を行うユーザーモデル同定手法について提案する。

2. ユーザーモデルの自動獲得手法

表現する分野(例:「人」「家族」「会社」等)を限定した自然語入力では限定しないときと比較して、入力文を形態素解析して得られる形態素間に強い意味的関係(例:形態素間の共起関係)の存在が認められる。例えば、「人」という分野での記述「得意料理がある女性」中に含まれる一般名詞「得意」「料理」「女性」を考える。「人」に対する意味的関係の強さをシソーラス上での接続関係の距離の近さと定義すると、「得意」よりも「女性」の方が「人」に対する意味的関係がより強いことがわかる。

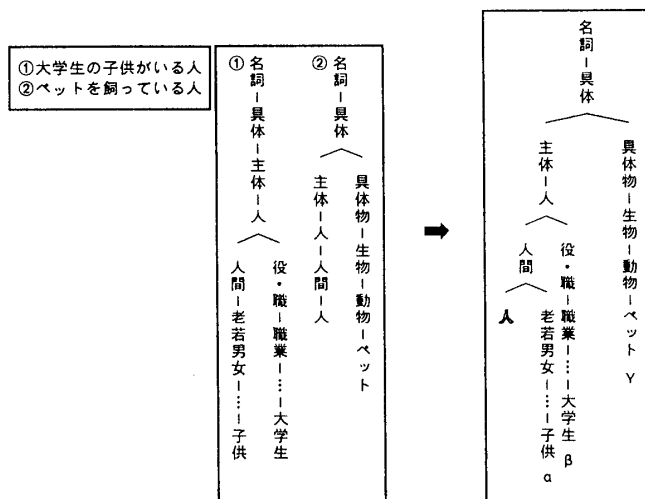
このことを利用して、分野限定(本稿では「人」に限定する)の自然語入力文からユーザーモデルを定義するために、形態素の中から意味属性を豊富に含む一般名詞を抽出しこれらを属性としたユーザーモデルの自動獲得を行う。以下に本手法のアルゴリズムを示す。

- ①「人」の属性を特徴づける文(定義文)に対して、形態素解析を行い一般名詞(用言性名詞も含む)を抽出する。
- ②汎用一般名詞シソーラス⁽¹⁾上での名詞間の接続関係の抽出を行い、「人」に対するシソーラス(分野依存シソーラス)の再構築を行う。
- ③獲得した意味属性、及び一般名詞間の接続関係距離の近さを得点化し蓄積する。
- ④定義文中から抽出された一般名詞の接続関係と得点化された距離の近さにより、ユーザーモデルを定義する。

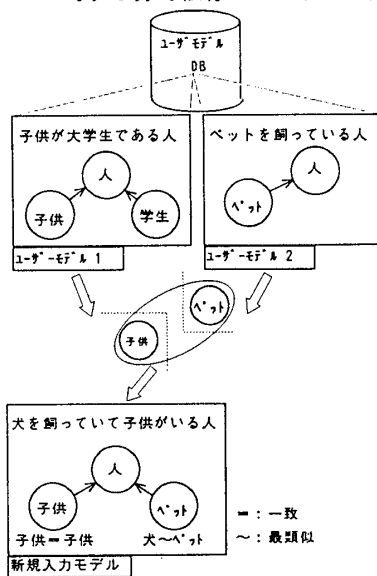
この結果、各定義文毎にその文が表現するユーザーモデルの獲得・蓄積が行われる(図1)。

3. ユーザーモデル同定手法

新規入力文が表現する未知のユーザーモデルと上記手法により獲得されたユーザーモデルとの間で、各文の一般名詞間の意味的関係の類似照合により、新規入力文が表現するユーザーモデルの同定を行う。ここで一般名詞間の類似照合とは、分野依存シソーラスにおいて基準となる名詞からの絶対距離の近さで照合を行うこととする。ユーザーモ



※ α、β、γは各名詞に与えられた得点
【図1 人に対する分野依存シソーラスの部分例】



【図2 新規モデルの同定】

デル同定手法のアルゴリズムは以下の通りである。

- ①新規入力文の形態素解析を行い、一般名詞を抽出する。
- ②①で得られた一般名詞に対して分野依存シソーラスに存在する一般名詞から意味的関係の一致した(一致しない場合は最も得点の近い)一般名詞を抽出する。
- ③類似名詞を包含するユーザーモデルを検索して、それらのインターセクションであるユーザーモデルを新規入力文のユーザーモデルとして同定する(図2)。

4. 商品販売セールストーク選択システムの試作

4.1 システム処理の流れ

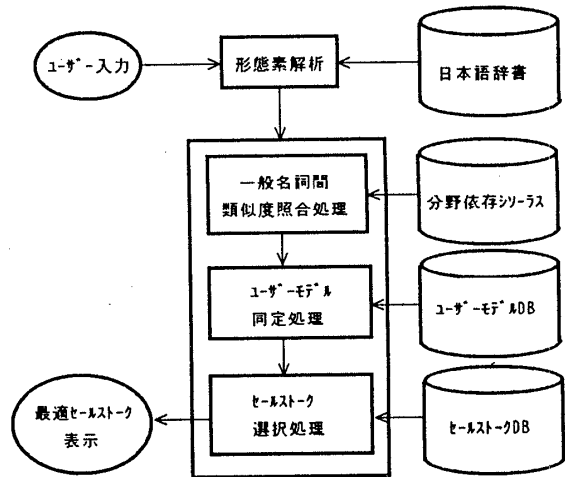
以上の手法を用いて、新規入力文が表現するユーザーモデルにふさわしい商品販売セールストーク選択システムを試作した。処理の流れを図3に示す。

まず商品販売対象である「人」に対する定義文100文をシステムに入力し、一般名詞シソーラスから分野依存シソーラスを構築しユーザーモデルの獲得・蓄積を行う。各ユーザーモデルには各モデルに適切なセールストークを付与しておく。このとき、定義文中の名詞の数により名詞個々のユーザーモデルに対する重要度が異なると仮定し、それぞれの名詞の各ユーザーモデルに対する重み付けを行う。これは、文中の名詞の数によるnormalized scoreの各セールストーク候補に対する付与により行う。例えば、図2中のユーザーモデル1に対しては、文中に3個の名詞を含むために各セールストークに対してnormalized scoreとして1/3が与えられる(図4)。そして、新規入力文に対して、その文が表現するユーザーモデルを同定し上記のnormalized scoreによりセールストークの優先順位を決定し適切なセールストークを表示する。

一例として、「犬を飼っていて、子供がいる人」という文の解析例を以下に示す。この新規文が入力されると、まず形態素解析により一般名詞「犬」「子供」「人」が抽出される。これらの各名詞に対して、分野依存シソーラスから一般名詞間類似度照合処理により各名詞に一致、或いは最類似名詞の自動抽出を行う。その結果、図2に示すように「犬〜ペット、子供=子供、人=人」と類似名詞が抽出され、ユーザーモデル同定処理により新規入力文が表現するユーザーモデルは、図2におけるモデル1とモデル2の各一部分の属性を持つユーザーモデルであることが同定される。最後にセールストーク選択処理がモデルに付与してあるセールストークの中から優先順位に従って表示する。上記の新規入力文では、各セールストーク候補に対する得点は図5のようになり、セールストーク5、2の順に優先順位がつけられ選択、表示される。

4.2 システムの評価

本システムにより、新規入力文の表現するユーザーモデルに対するユーザーモデルの同定について評価を行った。評価では、システムにより自動抽出されるユーザーモデルを構成する名詞と、人間が分野依存シソーラスから抽出する最適名詞候補(例:図6)とを比較したときの一般名詞間の照合成功率(図7)、及び新規入力文に対する有効セールストークの選択成功率を検証した。



【図3 商品販売セールストーク選択システムの処理の流れ】

この結果、100定義文による分野依存シソーラス上の第3候補までの一般名詞レベルでの照合成功率は75.0%であった。残りの25.0%中には、類似名詞が多数存在するために候補が絞りきれないものが存在する。これらに対しては、各一般名詞への意味属性の追加を行うことで類似名詞の照合成功率向上が期待できる。また、セールストークの選択では、本システムにより選択された第1位の候補に対して約80%の精度で有効なセールストークの選択ができ、本手法の有効性が示された。

5. まとめ

自然語入力文中に存在する一般名詞間の意味的関係の自動獲得を行うユーザーモデルの自動獲得手法、及び獲得・蓄積された意味的関係の類似照合により同定を行うユーザーモデル同定手法について提案し、上記手法を適用したセールストーク選択システムの試作とその評価を行った。その結果、未知のユーザーモデルに対する同定率は約80%となり本手法の有効性が示された。

<参考文献>

- (1) 宮崎他: 階層的単語属性を用いた同形語の自動読み分け法, 信学論, Vol. J68-D, No. 3, pp. 392-399 (1985)

定義文	セールストーク候補	normalized score
例A1 子供が大学生である人	1, 4, 5	1/3
例A2 犬を飼っている人	2, 5	1/2
例A3 部屋が多い家に住んでいる人	3, 4	1/3

※中抜き文字は名詞を示す

【図4 定義文とそれにリンクしたセールストーク】

セールストーク
①デザインで決めませんか?
②動物にいたずらされない工夫がしてあります。
③各部屋間で通信ができる電話機は?
④ホームテレホン各部屋に置いてはいかが?
⑤コードレスはコードがなくて便利です。

セールストーク	score	(sum)	order
①	1/3	1/3	3
②	1/2	1/2	2
④	1/3	1/3	3
⑤	1/3+1/2	5/6	1

【図5 セールストークの得点化】

新規文: 犬を飼っていて、子供がいる人	第1候補	第2候補	第3候補
犬	ペット	物	息子
子供	子供	息子	子

【図6 人間による候補名詞の例】

一般名詞レベルでの類似度照合成功率		
一致	第一候補で一致する	67.3%
	第二候補で一致する	5.8%
	第三候補で一致する	1.9%
		75.0%

【図7 一般名詞での照合成功率】